

Step 1: Data Preprocessing

1.1 Load and inspect the dataset:

```
import pandas as pd

# Load the dataset

df = pd.read_csv(r"C:\Users\Rajisha\Desktop\RAJISHA FINAL CAPSTONE PROJECT\Nutrical Dataset.csv")

# Display the first five rows of the dataset
df.head()
```

	Category	Item	Serving Size
0	Breakfast	Egg McMuffin	4.8 oz (136 g)
1	Breakfast	Egg White Delight	4.8 oz (135 g)
2	Breakfast	Sausage McMuffin	3.9 oz (111 g)
3	Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)
4	Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161 g)

	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated
0	120	13.0		20
1	70	8.0		12
2	200	23.0		35
3	250	28.0		43
4	210	23.0		35

	Saturated Fat (% Daily Value)	Trans Fat	Fat	...	Carbohydrates
0	25	0.0	...		31
1	15	0.0	...		30
2	42	0.0	...		29
3	52	0.0	...		30
4	42	0.0	...		30

	Carbohydrates (% Daily Value)	Dietary Fiber
--	-------------------------------	---------------

0	10	4
1	10	4
2	10	4
3	10	4
4	10	4

Dietary Fiber (% Daily Value) Sugars Protein Vitamin A (% Daily Value) \

0	17	3	17
10			
1	17	3	18
6			
2	17	2	14
8			
3	17	2	21
15			
4	17	2	21
6			

Vitamin C (% Daily Value) Calcium (% Daily Value) Iron (% Daily Value)

0	0	25
15		
1	0	25
8		
2	0	25
10		
3	0	30
15		
4	0	25
10		

[5 rows x 24 columns]

`print(df.head())`

Category		Item	Serving Size
Calories \			
0	Breakfast	Egg McMuffin	4.8 oz (136 g)
300			
1	Breakfast	Egg White Delight	4.8 oz (135 g)
250			
2	Breakfast	Sausage McMuffin	3.9 oz (111 g)
370			
3	Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)
450			
4	Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161 g)
400			

Calories from Fat Total Fat Total Fat (% Daily Value) Saturated

Fat \					
0	120	13.0			20
5.0					
1	70	8.0			12
3.0					
2	200	23.0			35
8.0					
3	250	28.0			43
10.0					
4	210	23.0			35
8.0					
Saturated Fat (% Daily Value) Trans Fat ... Carbohydrates \					
0	25	0.0	...		31
1	15	0.0	...		30
2	42	0.0	...		29
3	52	0.0	...		30
4	42	0.0	...		30
Carbohydrates (% Daily Value) Dietary Fiber \					
0	10			4	
1	10			4	
2	10			4	
3	10			4	
4	10			4	
Dietary Fiber (% Daily Value) Sugars Protein Vitamin A (% Daily Value) \					
0	17	3		17	
10					
1	17	3		18	
6					
2	17	2		14	
8					
3	17	2		21	
15					
4	17	2		21	
6					
Vitamin C (% Daily Value) Calcium (% Daily Value) Iron (% Daily Value)					
0	0			25	
15					
1	0			25	
8					
2	0			25	
10					
3	0			30	
15					
4	0			25	

```
10
```

```
[5 rows x 24 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 260 entries, 0 to 259
```

```
Data columns (total 24 columns):
```

#	Column	Non-Null Count	Dtype
0	Category	260 non-null	object
1	Item	260 non-null	object
2	Serving Size	260 non-null	object
3	Calories	260 non-null	int64
4	Calories from Fat	260 non-null	int64
5	Total Fat	260 non-null	float64
6	Total Fat (% Daily Value)	260 non-null	int64
7	Saturated Fat	260 non-null	float64
8	Saturated Fat (% Daily Value)	260 non-null	int64
9	Trans Fat	260 non-null	float64
10	Cholesterol	260 non-null	int64
11	Cholesterol (% Daily Value)	260 non-null	int64
12	Sodium	260 non-null	int64
13	Sodium (% Daily Value)	260 non-null	int64
14	Carbohydrates	260 non-null	int64
15	Carbohydrates (% Daily Value)	260 non-null	int64
16	Dietary Fiber	260 non-null	int64
17	Dietary Fiber (% Daily Value)	260 non-null	int64
18	Sugars	260 non-null	int64
19	Protein	260 non-null	int64
20	Vitamin A (% Daily Value)	260 non-null	int64
21	Vitamin C (% Daily Value)	260 non-null	int64
22	Calcium (% Daily Value)	260 non-null	int64
23	Iron (% Daily Value)	260 non-null	int64

```
dtypes: float64(3), int64(18), object(3)
```

```
memory usage: 48.9+ KB
```

```
len(df)
```

```
260
```

```
df.columns
```

```
Index(['Category', 'Item', 'Serving Size', 'Calories', 'Calories from Fat',
```

```
      'Total Fat', 'Total Fat (% Daily Value)', 'Saturated Fat',  
      'Saturated Fat (% Daily Value)', 'Trans Fat', 'Cholesterol',  
      'Cholesterol (% Daily Value)', 'Sodium', 'Sodium (% Daily Value)',  
      'Carbohydrates', 'Carbohydrates (% Daily Value)', 'Dietary
```

```
Fiber',
      'Dietary Fiber (% Daily Value)', 'Sugars', 'Protein',
      'Vitamin A (% Daily Value)', 'Vitamin C (% Daily Value)',
      'Calcium (% Daily Value)', 'Iron (% Daily Value)'],
      dtype='object')

len(df.columns)

24
```

1.2 Handle missing values and data cleaning:

```
# Check for missing values
df.isnull().sum()

Category                                0
Item                                    0
Serving Size                            0
Calories                                0
Calories from Fat                       0
Total Fat                               0
Total Fat (% Daily Value)               0
Saturated Fat                           0
Saturated Fat (% Daily Value)           0
Trans Fat                               0
Cholesterol                             0
Cholesterol (% Daily Value)             0
Sodium                                  0
Sodium (% Daily Value)                  0
Carbohydrates                           0
Carbohydrates (% Daily Value)           0
Dietary Fiber                           0
Dietary Fiber (% Daily Value)           0
Sugars                                  0
Protein                                 0
Vitamin A (% Daily Value)               0
Vitamin C (% Daily Value)               0
Calcium (% Daily Value)                 0
Iron (% Daily Value)                    0
dtype: int64
```

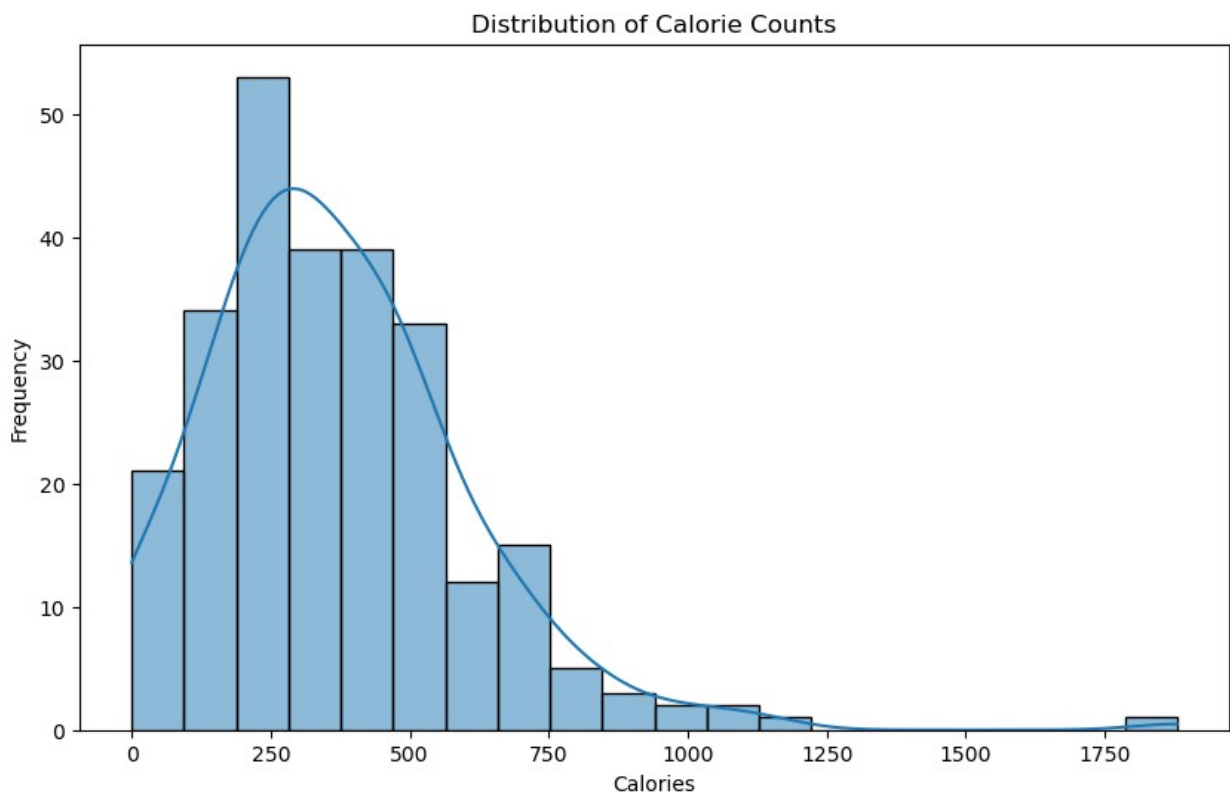
- As we can see there are no missing values in this dataset, we are good to proceed further.

Step 2: Exploratory Data Analysis (EDA)

2.1 Analyze the distribution of calorie counts:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Distribution of calorie counts
plt.figure(figsize=(10, 6))
sns.histplot(df['Calories'], bins=20, kde=True)
plt.title('Distribution of Calorie Counts')
plt.xlabel('Calories')
plt.ylabel('Frequency')
plt.show()
```

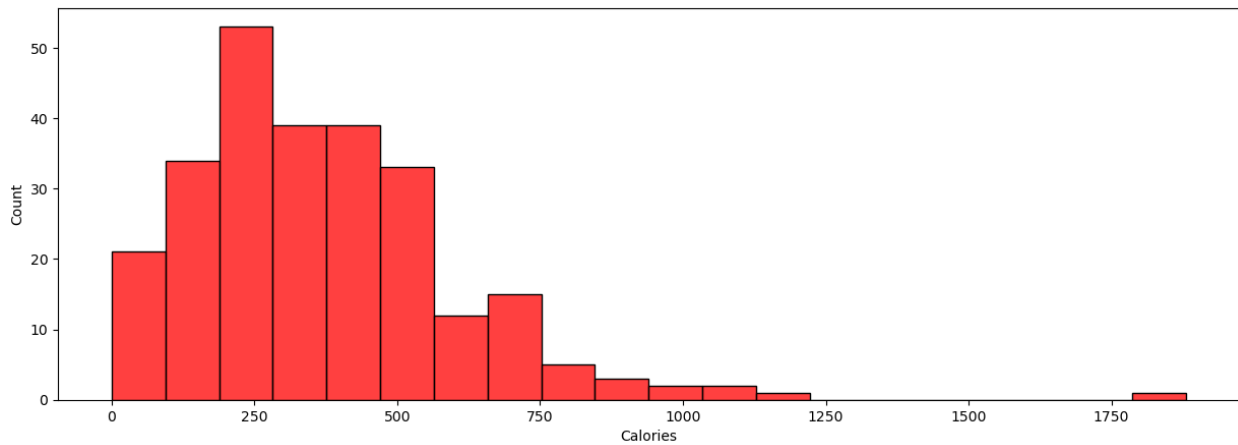


```
import matplotlib.pyplot as plt
import seaborn as sns

# Create a figure and axis object
fig, ax = plt.subplots(figsize=(15, 5))

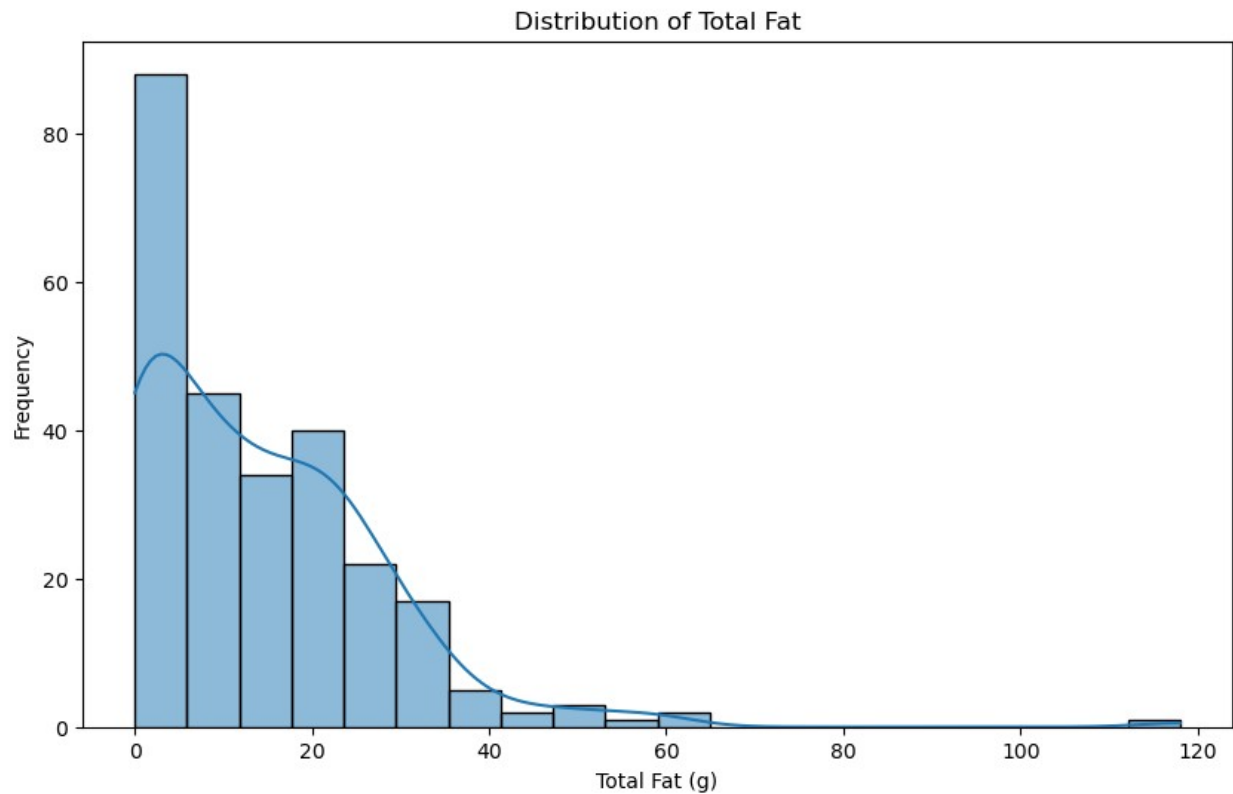
# Create a histogram using seaborn's histplot function
sns.histplot(x='Calories', data=df, ax=ax, color='red', bins=20)
```

```
# Show the plot  
plt.show()
```

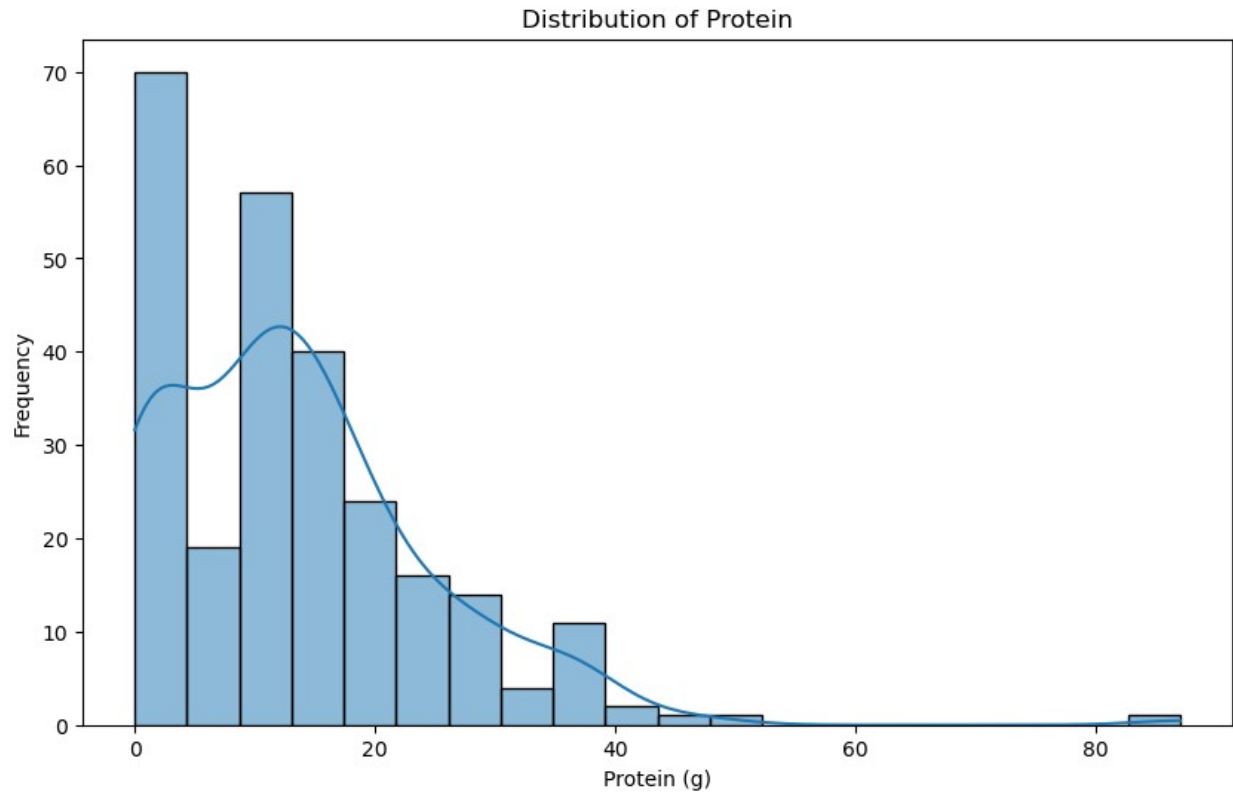


2.2 Explore the nutritional content:

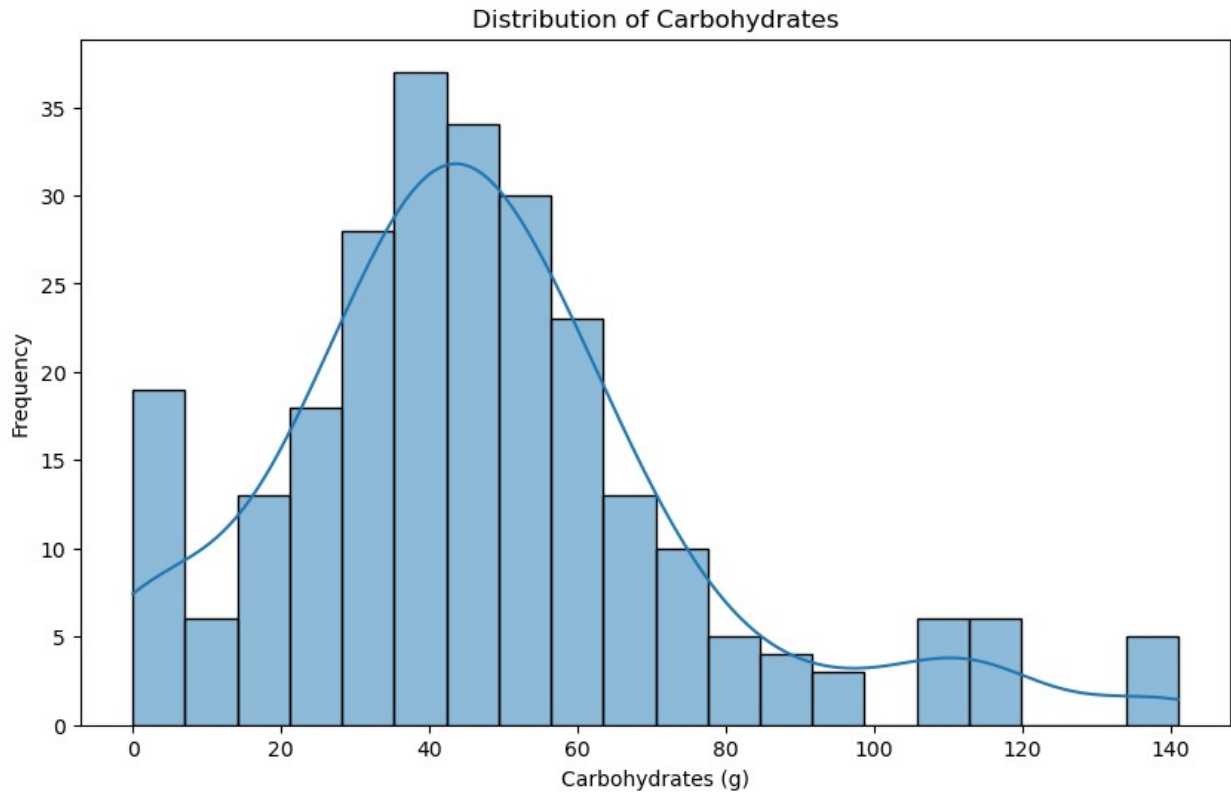
```
# Distribution of fat content  
plt.figure(figsize=(10, 6))  
sns.histplot(df['Total Fat'], bins=20, kde=True)  
plt.title('Distribution of Total Fat')  
plt.xlabel('Total Fat (g)')  
plt.ylabel('Frequency')  
plt.show()
```



```
# Distribution of Protein content
plt.figure(figsize=(10, 6))
sns.histplot(df['Protein'], bins=20, kde=True)
plt.title('Distribution of Protein')
plt.xlabel('Protein (g)')
plt.ylabel('Frequency')
plt.show()
```

```
# Distribution of Carbohydrates content
plt.figure(figsize=(10, 6))
sns.histplot(df['Carbohydrates'], bins=20, kde=True)
plt.title('Distribution of Carbohydrates')
plt.xlabel('Carbohydrates (g)')
plt.ylabel('Frequency')
plt.show()
```

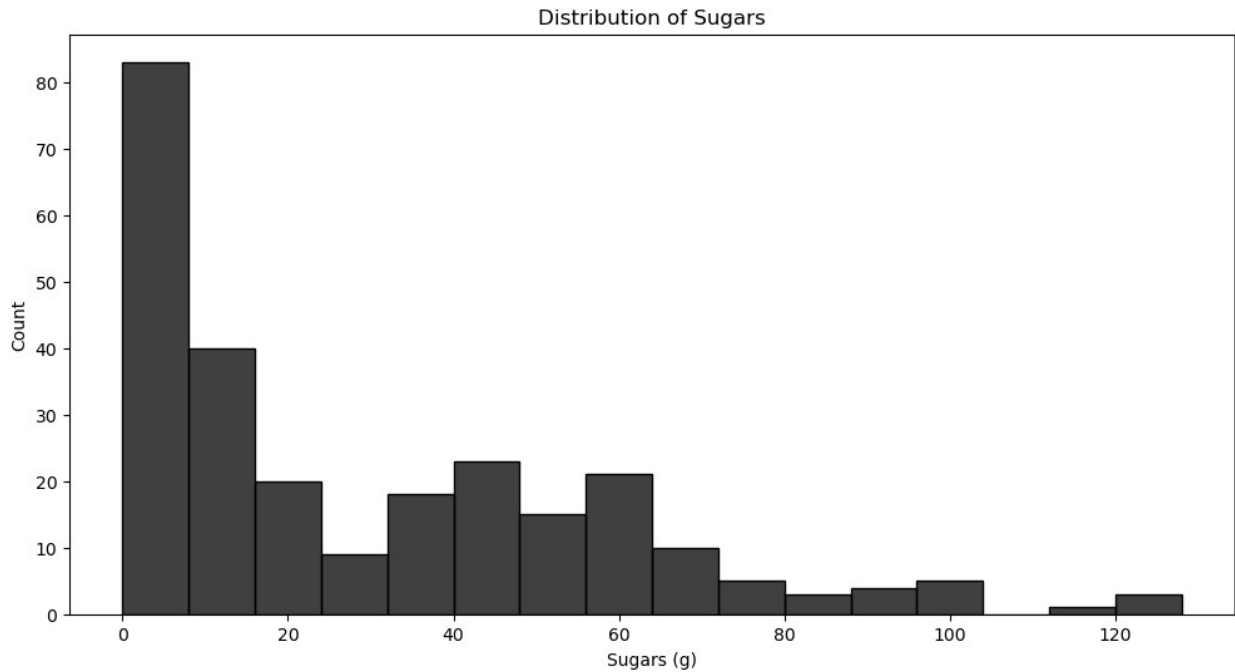


```
import matplotlib.pyplot as plt
import seaborn as sns

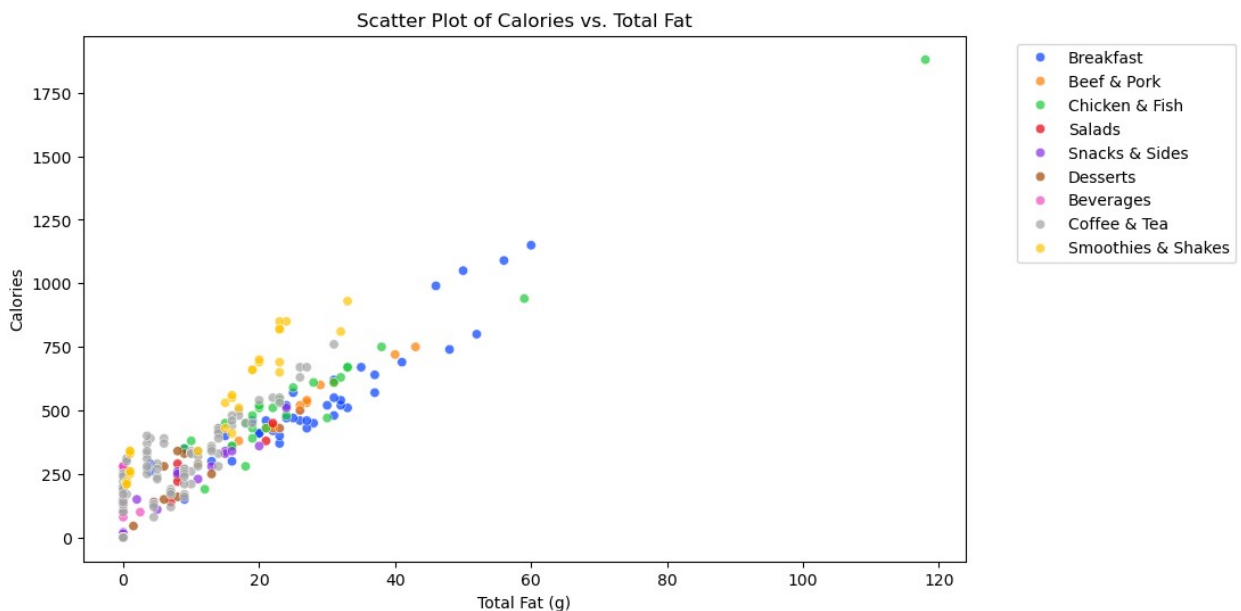
# Create a figure and axis object
fig, ax = plt.subplots(figsize=(12, 6))

# Create a histogram using seaborn's histplot function
sns.histplot(x='Sugars', data=df, ax=ax, color='black', bins=16)

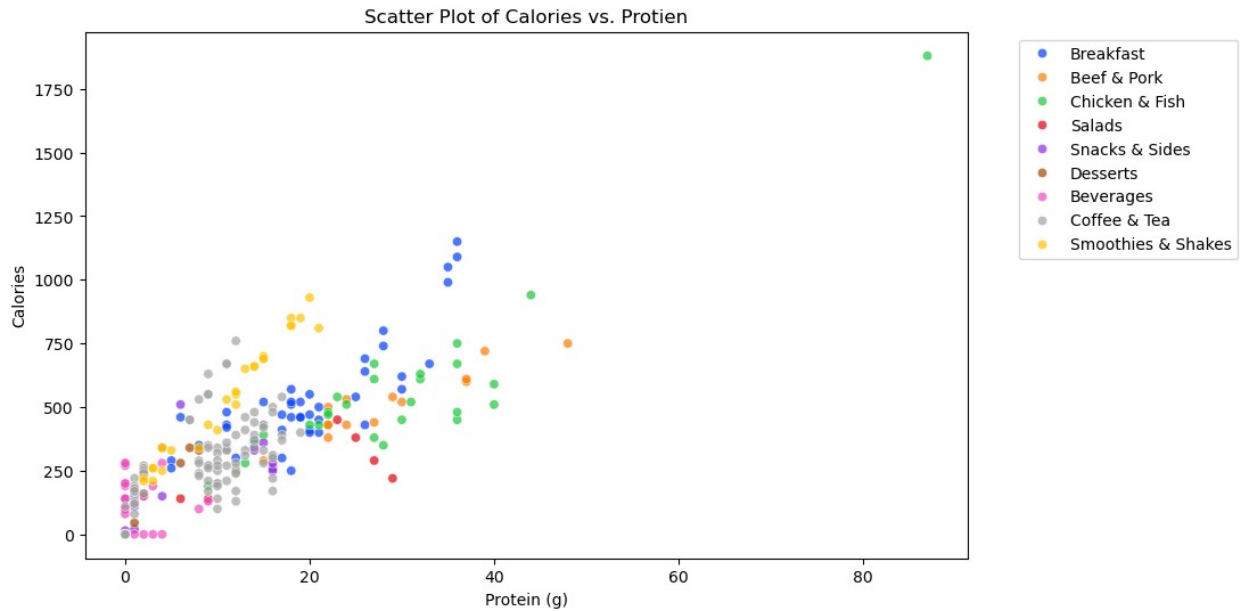
ax.set_title('Distribution of Sugars')
ax.set_xlabel('Sugars (g)')
ax.set_ylabel('Count')
plt.show()
```



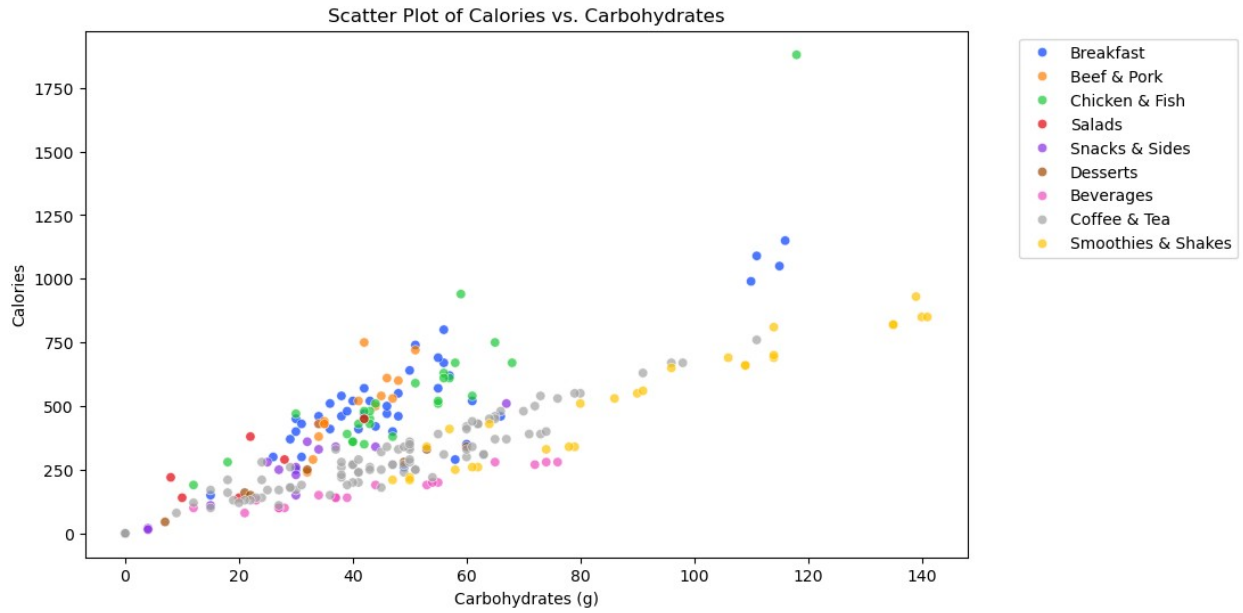
```
# Scatter Plot of Calories vs. Total Fat
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Total Fat', y='Calories', data=df, hue='Category',
               palette='bright', alpha=0.7)
plt.title('Scatter Plot of Calories vs. Total Fat')
plt.xlabel('Total Fat (g)')
plt.ylabel('Calories')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```



```
# Scatter Plot of Calories vs. Protein
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Protein', y='Calories', data=df, hue='Category',
palette='bright', alpha=0.7)
plt.title('Scatter Plot of Calories vs. Protein')
plt.xlabel('Protein (g)')
plt.ylabel('Calories')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```



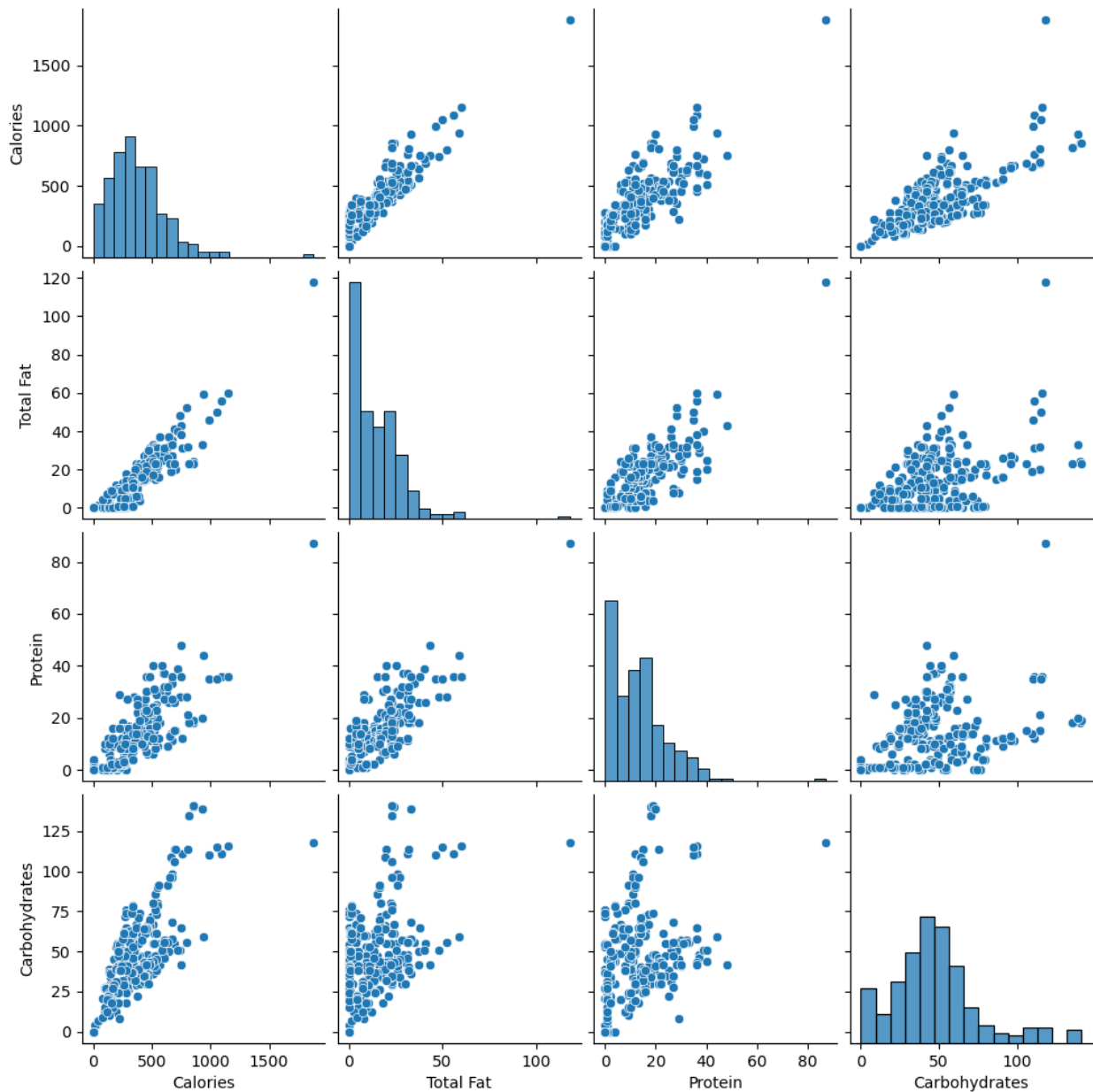
```
# Scatter Plot of Calories vs. Carbohydrate
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Carbohydrates', y='Calories', data=df,
hue='Category', palette='bright', alpha=0.7)
plt.title('Scatter Plot of Calories vs. Carbohydrates')
plt.xlabel('Carbohydrates (g)')
plt.ylabel('Calories')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```



2.3 Identify trends and patterns:

Pairplot to identify relationships between different nutritional components

```
sns.pairplot(df[['Calories', 'Total Fat', 'Protein',  
'Carbohydrates']])  
plt.show()
```



2.4. Correlation Matrix

```
df.corr(numeric_only=True).round(2)
```

	Calories	Calories from Fat	Total Fat
Calories	1.00	0.90	0.90
Calories from Fat	0.90	1.00	1.00
Total Fat	0.90	1.00	1.00
Total Fat (% Daily Value)	0.90	1.00	1.00

Saturated Fat	0.85	0.85	0.85
Saturated Fat (% Daily Value)	0.85	0.85	0.85
Trans Fat	0.52	0.43	0.43
Cholesterol	0.60	0.68	0.68
Cholesterol (% Daily Value)	0.60	0.68	0.68
Sodium	0.71	0.85	0.85
Sodium (% Daily Value)	0.71	0.85	0.85
Carbohydrates	0.78	0.46	0.46
Carbohydrates (% Daily Value)	0.78	0.46	0.46
Dietary Fiber	0.54	0.58	0.58
Dietary Fiber (% Daily Value)	0.54	0.58	0.58
Sugars	0.26	-0.12	-0.12
Protein	0.79	0.81	0.81
Vitamin A (% Daily Value)	0.11	0.06	0.05
Vitamin C (% Daily Value)	-0.07	-0.09	-0.09
Calcium (% Daily Value)	0.43	0.16	0.16
Iron (% Daily Value)	0.64	0.74	0.73
		Total Fat (% Daily Value)	Saturated
Fat \			
Calories			0.90
0.85			
Calories from Fat			1.00
0.85			
Total Fat			1.00
0.85			
Total Fat (% Daily Value)			1.00
0.85			
Saturated Fat			0.85
1.00			
Saturated Fat (% Daily Value)			0.85
1.00			
Trans Fat			0.43

0.62		
Cholesterol	0.68	
0.63		
Cholesterol (% Daily Value)	0.68	
0.63		
Sodium	0.85	
0.58		
Sodium (% Daily Value)	0.85	
0.59		
Carbohydrates	0.46	
0.59		
Carbohydrates (% Daily Value)	0.46	
0.59		
Dietary Fiber	0.58	
0.35		
Dietary Fiber (% Daily Value)	0.58	
0.35		
Sugars	-0.12	
0.20		
Protein	0.81	
0.60		
Vitamin A (% Daily Value)	0.05	
0.06		
Vitamin C (% Daily Value)	-0.09	-
0.18		
Calcium (% Daily Value)	0.16	
0.40		
Iron (% Daily Value)	0.74	
0.58		

	Saturated Fat (% Daily Value)	Trans
Fat \		
Calories	0.85	
0.52		
Calories from Fat	0.85	
0.43		
Total Fat	0.85	
0.43		
Total Fat (% Daily Value)	0.85	
0.43		
Saturated Fat	1.00	
0.62		
Saturated Fat (% Daily Value)	1.00	
0.62		
Trans Fat	0.62	
1.00		
Cholesterol	0.63	
0.25		
Cholesterol (% Daily Value)	0.63	

0.25		
Sodium	0.59	
0.19		
Sodium (% Daily Value)	0.59	
0.19		
Carbohydrates	0.59	
0.46		
Carbohydrates (% Daily Value)	0.59	
0.46		
Dietary Fiber	0.36	
0.05		
Dietary Fiber (% Daily Value)	0.35	
0.06		
Sugars	0.20	
0.33		
Protein	0.61	
0.39		
Vitamin A (% Daily Value)	0.07	
0.08		
Vitamin C (% Daily Value)	-0.18	-
0.08		
Calcium (% Daily Value)	0.40	
0.39		
Iron (% Daily Value)	0.58	
0.33		

	Cholesterol	Cholesterol (% Daily
Value) \		
Calories	0.60	
0.60		
Calories from Fat	0.68	
0.68		
Total Fat	0.68	
0.68		
Total Fat (% Daily Value)	0.68	
0.68		
Saturated Fat	0.63	
0.63		
Saturated Fat (% Daily Value)	0.63	
0.63		
Trans Fat	0.25	
0.25		
Cholesterol	1.00	
1.00		
Cholesterol (% Daily Value)	1.00	
1.00		
Sodium	0.62	
0.62		
Sodium (% Daily Value)	0.62	

0.62		
Carbohydrates	0.27	
0.27		
Carbohydrates (% Daily Value)	0.27	
0.27		
Dietary Fiber	0.44	
0.43		
Dietary Fiber (% Daily Value)	0.44	
0.44		
Sugars	-0.14	-
0.14		
Protein	0.56	
0.56		
Vitamin A (% Daily Value)	0.08	
0.08		
Vitamin C (% Daily Value)	-0.08	-
0.08		
Calcium (% Daily Value)	0.13	
0.13		
Iron (% Daily Value)	0.65	
0.65		

	Sodium	...	Carbohydrates \
Calories	0.71	...	0.78
Calories from Fat	0.85	...	0.46
Total Fat	0.85	...	0.46
Total Fat (% Daily Value)	0.85	...	0.46
Saturated Fat	0.58	...	0.59
Saturated Fat (% Daily Value)	0.59	...	0.59
Trans Fat	0.19	...	0.46
Cholesterol	0.62	...	0.27
Cholesterol (% Daily Value)	0.62	...	0.27
Sodium	1.00	...	0.20
Sodium (% Daily Value)	1.00	...	0.20
Carbohydrates	0.20	...	1.00
Carbohydrates (% Daily Value)	0.20	...	1.00
Dietary Fiber	0.69	...	0.22
Dietary Fiber (% Daily Value)	0.69	...	0.23
Sugars	-0.43	...	0.76
Protein	0.87	...	0.35
Vitamin A (% Daily Value)	0.08	...	0.08
Vitamin C (% Daily Value)	-0.03	...	-0.03
Calcium (% Daily Value)	-0.02	...	0.59
Iron (% Daily Value)	0.87	...	0.21

	Carbohydrates (% Daily Value)	Dietary
Fiber \		
Calories		0.78
0.54		

Calories from Fat	0.46	
0.58		
Total Fat	0.46	
0.58		
Total Fat (% Daily Value)	0.46	
0.58		
Saturated Fat	0.59	
0.35		
Saturated Fat (% Daily Value)	0.59	
0.36		
Trans Fat	0.46	
0.05		
Cholesterol	0.27	
0.44		
Cholesterol (% Daily Value)	0.27	
0.43		
Sodium	0.20	
0.69		
Sodium (% Daily Value)	0.20	
0.69		
Carbohydrates	1.00	
0.22		
Carbohydrates (% Daily Value)	1.00	
0.22		
Dietary Fiber	0.22	
1.00		
Dietary Fiber (% Daily Value)	0.23	
0.99		
Sugars	0.76	
-0.30		
Protein	0.35	
0.64		
Vitamin A (% Daily Value)	0.08	
0.34		
Vitamin C (% Daily Value)	-0.04	
0.14		
Calcium (% Daily Value)	0.59	
0.03		
Iron (% Daily Value)	0.21	
0.74		
Dietary Fiber (% Daily Value)		Sugars
Protein \		
Calories	0.54	0.26
0.79		
Calories from Fat	0.58	-0.12
0.81		
Total Fat	0.58	-0.12
0.81		

Total Fat (% Daily Value)	0.58	-0.12
0.81		
Saturated Fat	0.35	0.20
0.60		
Saturated Fat (% Daily Value)	0.35	0.20
0.61		
Trans Fat	0.06	0.33
0.39		
Cholesterol	0.44	-0.14
0.56		
Cholesterol (% Daily Value)	0.44	-0.14
0.56		
Sodium	0.69	-0.43
0.87		
Sodium (% Daily Value)	0.69	-0.42
0.87		
Carbohydrates	0.23	0.76
0.35		
Carbohydrates (% Daily Value)	0.23	0.76
0.35		
Dietary Fiber	0.99	-0.30
0.64		
Dietary Fiber (% Daily Value)	1.00	-0.29
0.66		
Sugars	-0.29	1.00
-0.18		
Protein	0.66	-0.18
1.00		
Vitamin A (% Daily Value)	0.36	0.05
0.21		
Vitamin C (% Daily Value)	0.15	-0.07
-0.05		
Calcium (% Daily Value)	0.05	0.60
0.33		
Iron (% Daily Value)	0.74	-0.36
0.79		

	Vitamin A (% Daily Value)	\
Calories	0.11	
Calories from Fat	0.06	
Total Fat	0.05	
Total Fat (% Daily Value)	0.05	
Saturated Fat	0.06	
Saturated Fat (% Daily Value)	0.07	
Trans Fat	0.08	
Cholesterol	0.08	
Cholesterol (% Daily Value)	0.08	
Sodium	0.08	
Sodium (% Daily Value)	0.08	

Carbohydrates	0.08
Carbohydrates (% Daily Value)	0.08
Dietary Fiber	0.34
Dietary Fiber (% Daily Value)	0.36
Sugars	0.05
Protein	0.21
Vitamin A (% Daily Value)	1.00
Vitamin C (% Daily Value)	0.07
Calcium (% Daily Value)	0.18
Iron (% Daily Value)	0.14

	Vitamin C (% Daily Value)	\
Calories	-0.07	
Calories from Fat	-0.09	
Total Fat	-0.09	
Total Fat (% Daily Value)	-0.09	
Saturated Fat	-0.18	
Saturated Fat (% Daily Value)	-0.18	
Trans Fat	-0.08	
Cholesterol	-0.08	
Cholesterol (% Daily Value)	-0.08	
Sodium	-0.03	
Sodium (% Daily Value)	-0.03	
Carbohydrates	-0.03	
Carbohydrates (% Daily Value)	-0.04	
Dietary Fiber	0.14	
Dietary Fiber (% Daily Value)	0.15	
Sugars	-0.07	
Protein	-0.05	
Vitamin A (% Daily Value)	0.07	
Vitamin C (% Daily Value)	1.00	
Calcium (% Daily Value)	-0.22	
Iron (% Daily Value)	0.00	

	Calcium (% Daily Value)	Iron (% Daily Value)
Calories	0.43	
0.64		
Calories from Fat	0.16	
0.74		
Total Fat	0.16	
0.73		
Total Fat (% Daily Value)	0.16	
0.74		
Saturated Fat	0.40	
0.58		
Saturated Fat (% Daily Value)	0.40	
0.58		
Trans Fat	0.39	

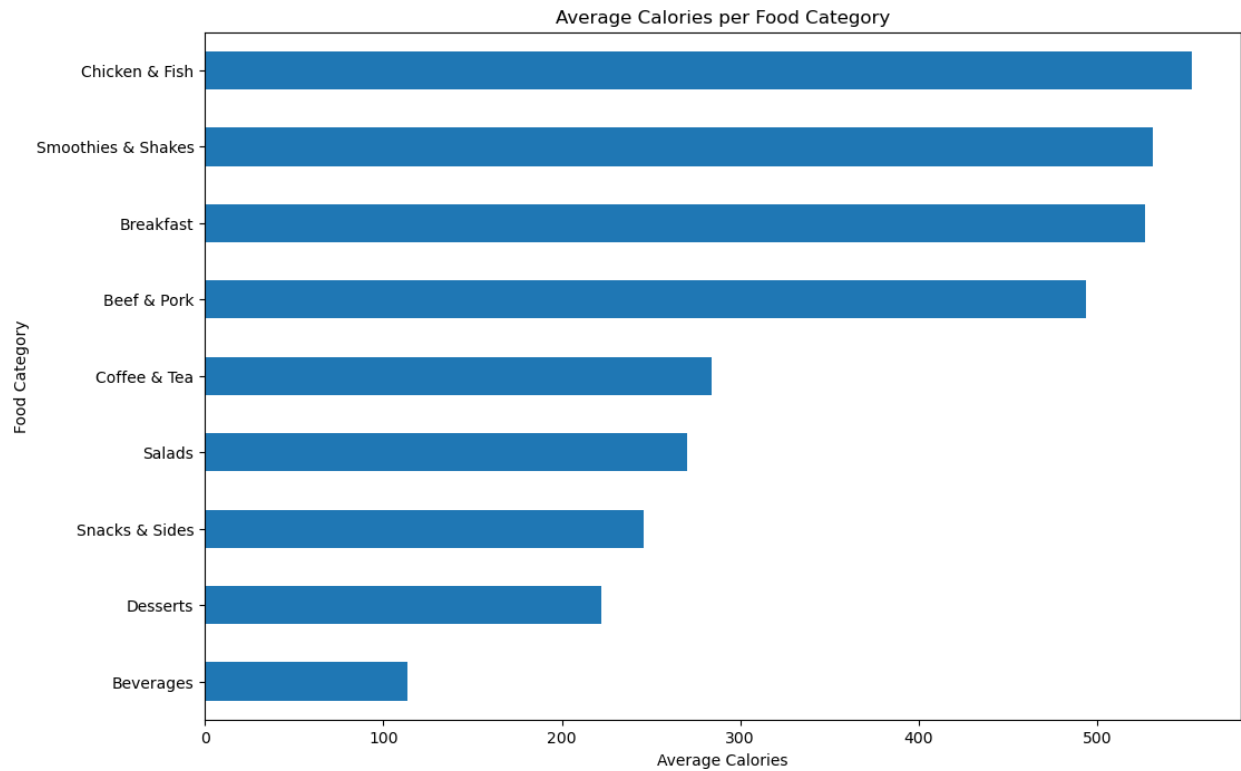
0.33	
Cholesterol	0.13
0.65	
Cholesterol (% Daily Value)	0.13
0.65	
Sodium	-0.02
0.87	
Sodium (% Daily Value)	-0.02
0.87	
Carbohydrates	0.59
0.21	
Carbohydrates (% Daily Value)	0.59
0.21	
Dietary Fiber	0.03
0.74	
Dietary Fiber (% Daily Value)	0.05
0.74	
Sugars	0.60
-0.36	
Protein	0.33
0.79	
Vitamin A (% Daily Value)	0.18
0.14	
Vitamin C (% Daily Value)	-0.22
0.00	
Calcium (% Daily Value)	1.00
0.03	
Iron (% Daily Value)	0.03
1.00	

[21 rows x 21 columns]

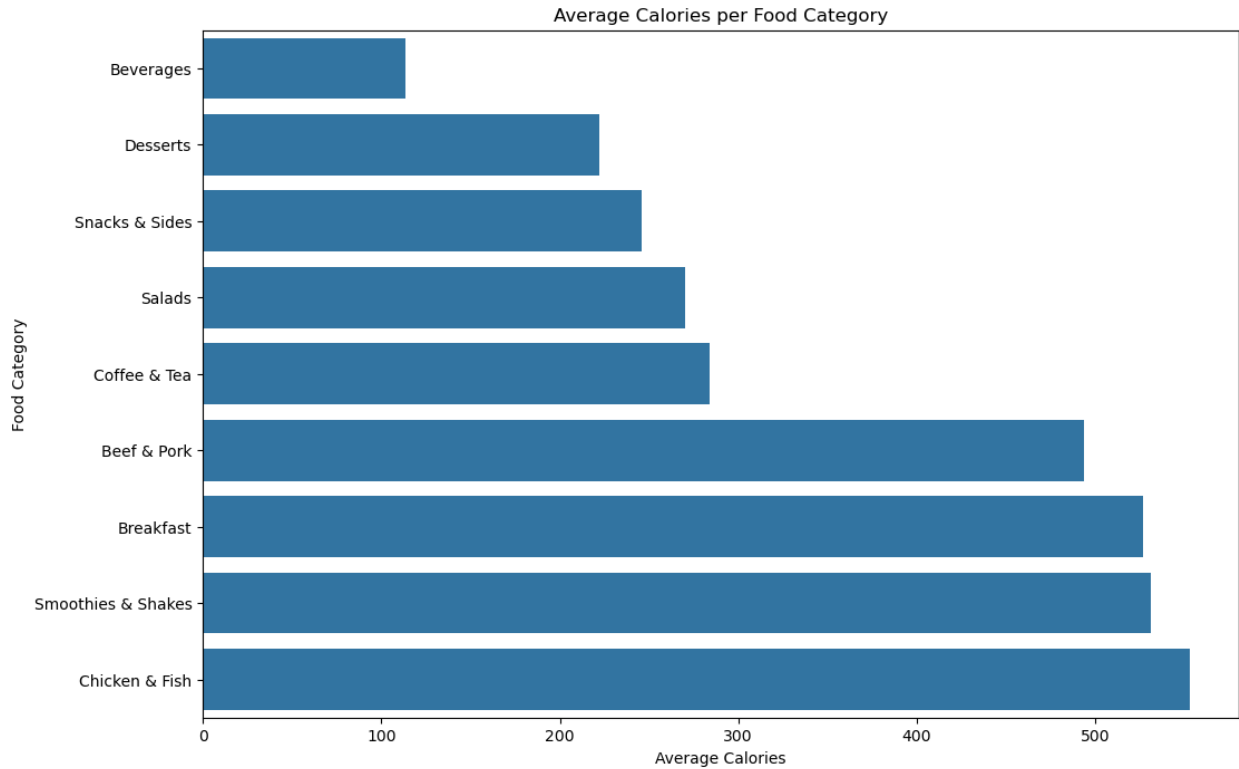
Step 3: Data Visualization

3.1 Create bar charts, histograms, and box plots:

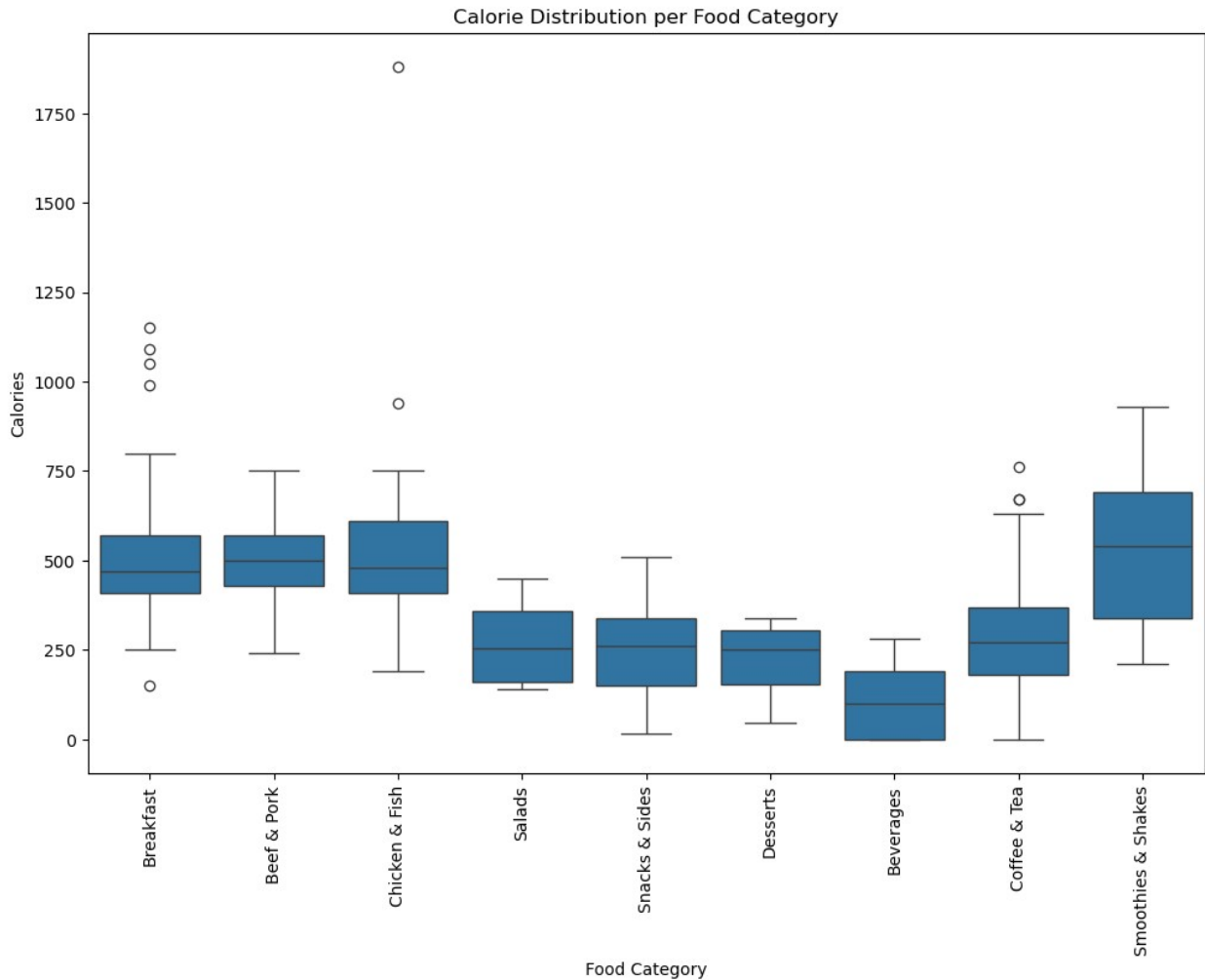
```
# Bar chart for average calories per food category
avg_calories_per_category = df.groupby('Category')
['Calories'].mean().sort_values()
plt.figure(figsize=(12, 8))
avg_calories_per_category.plot(kind='barh')
plt.title('Average Calories per Food Category')
plt.xlabel('Average Calories')
plt.ylabel('Food Category')
plt.show()
```



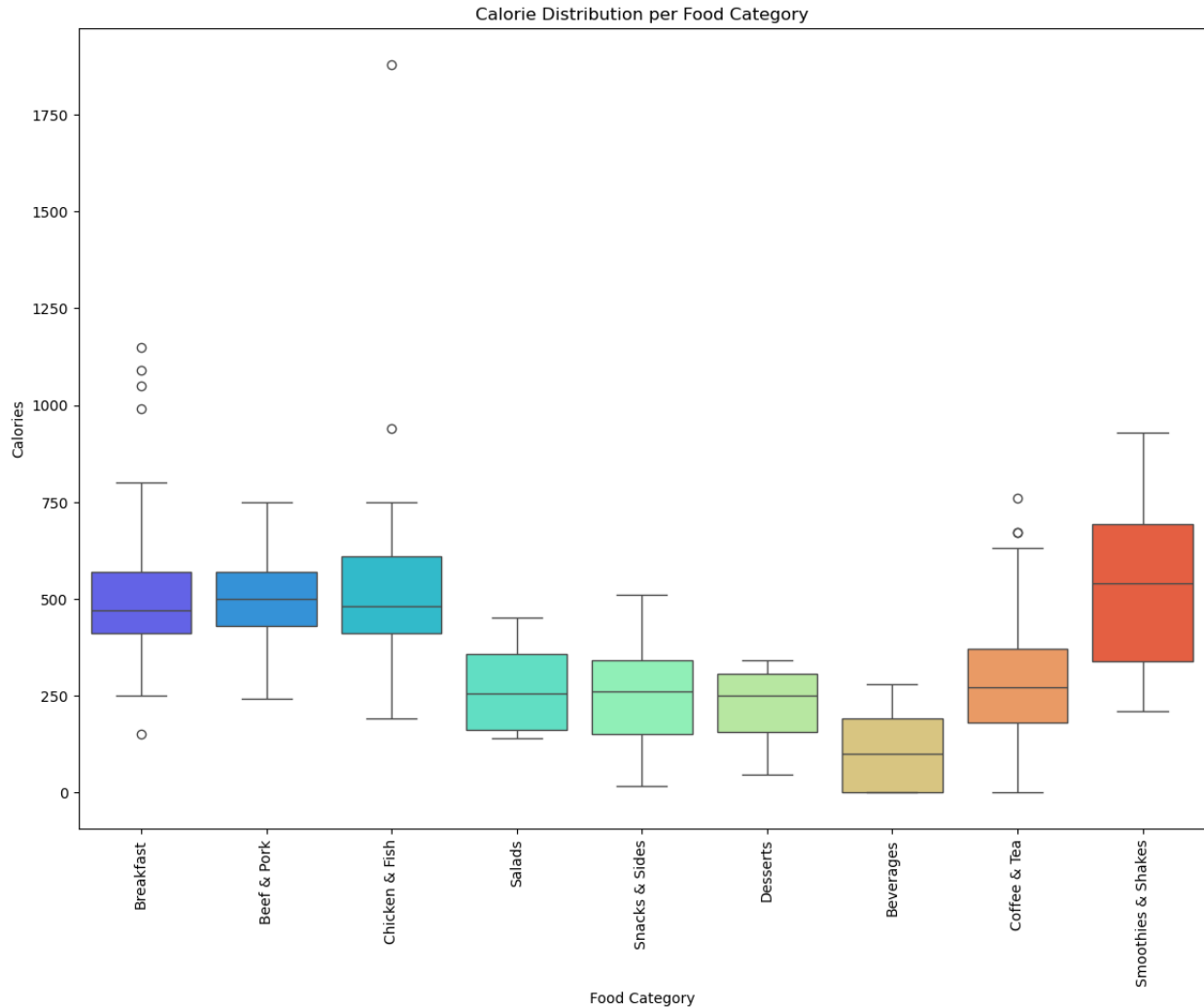
```
# Plot the bar chart
plt.figure(figsize=(12, 8))
sns.barplot(x=avg_calories_per_category.values,
            y=avg_calories_per_category.index)
plt.title('Average Calories per Food Category')
plt.xlabel('Average Calories')
plt.ylabel('Food Category')
plt.show()
```



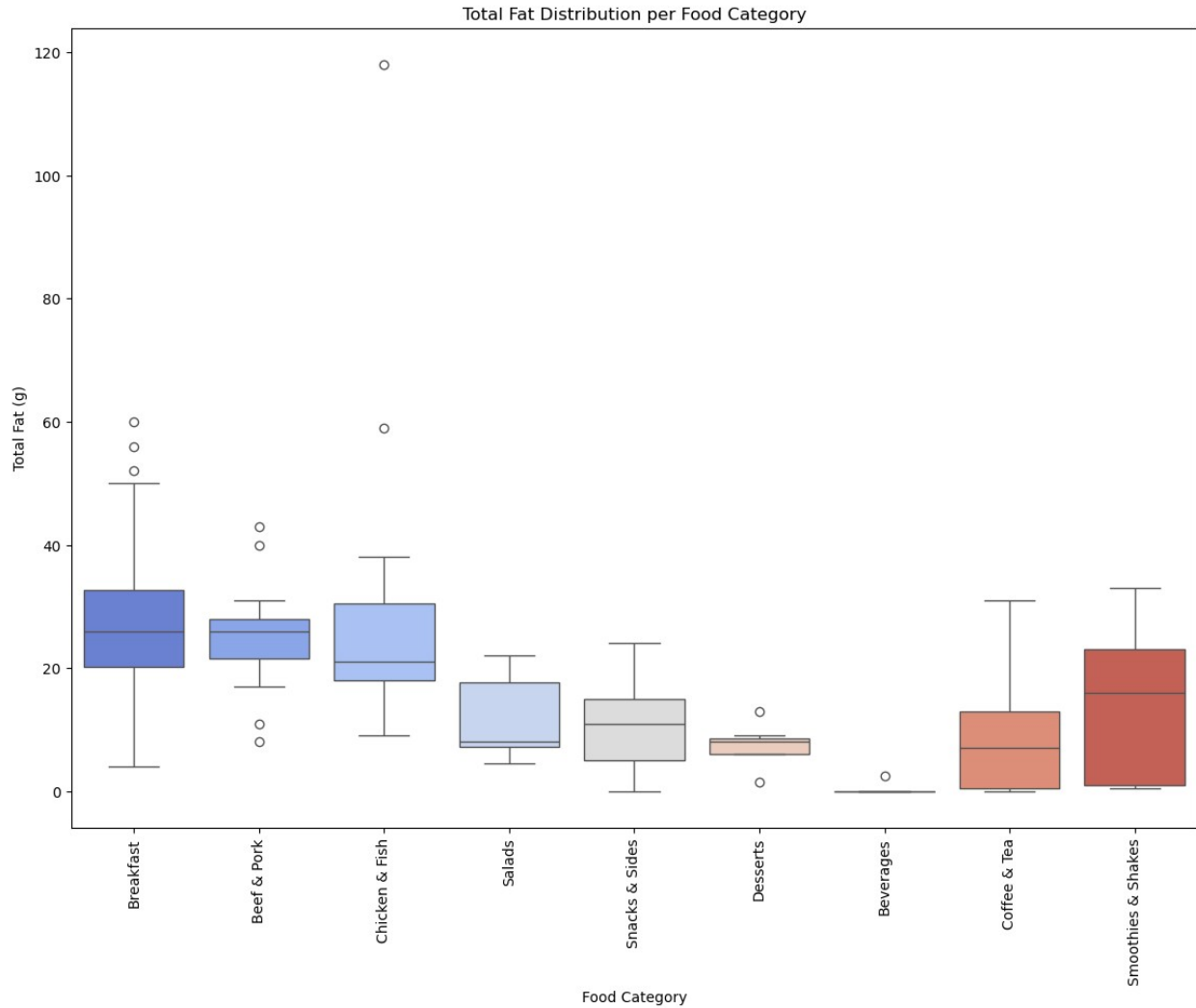
```
# Box plot for calorie distribution per category
plt.figure(figsize=(12, 8))
sns.boxplot(x='Category', y='Calories', data=df)
plt.xticks(rotation=90)
plt.title('Calorie Distribution per Food Category')
plt.xlabel('Food Category')
plt.ylabel('Calories')
plt.show()
```

```
# Box plot for calorie distribution per category
plt.figure(figsize=(14, 10))
sns.boxplot(x='Category', y='Calories', data=df, hue='Category',
palette='rainbow', dodge=False, legend=False)
plt.xticks(rotation=90)
plt.title('Calorie Distribution per Food Category')
plt.xlabel('Food Category')
plt.ylabel('Calories')
plt.show()
```



```
# Box plot for total fat distribution per category
plt.figure(figsize=(14, 10))
sns.boxplot(x='Category', y='Total Fat', data=df, hue='Category',
palette='coolwarm', dodge=False, legend=False)
plt.xticks(rotation=90)
plt.title('Total Fat Distribution per Food Category')
plt.xlabel('Food Category')
plt.ylabel('Total Fat (g)')
plt.show()
```

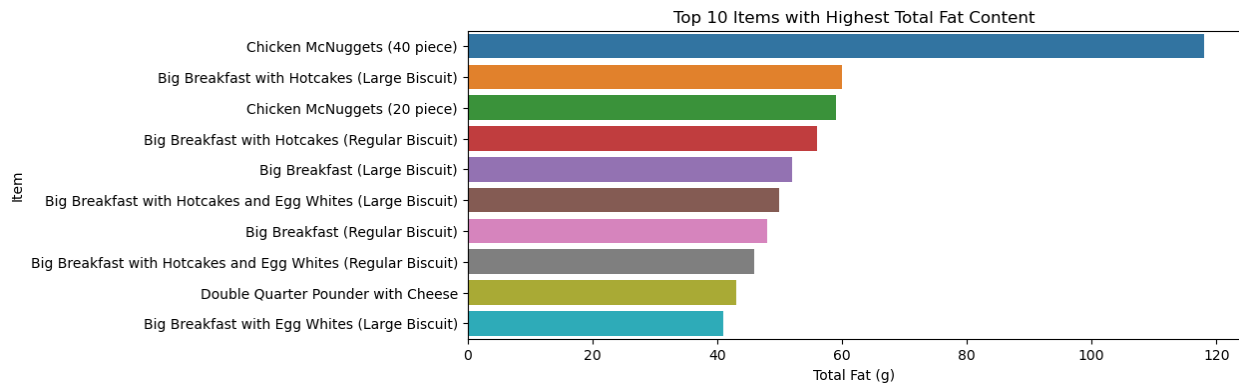


```
# Top 10 Items with Highest Total Fat Content
max_fat = df.sort_values('Total Fat', ascending=False).head(10)

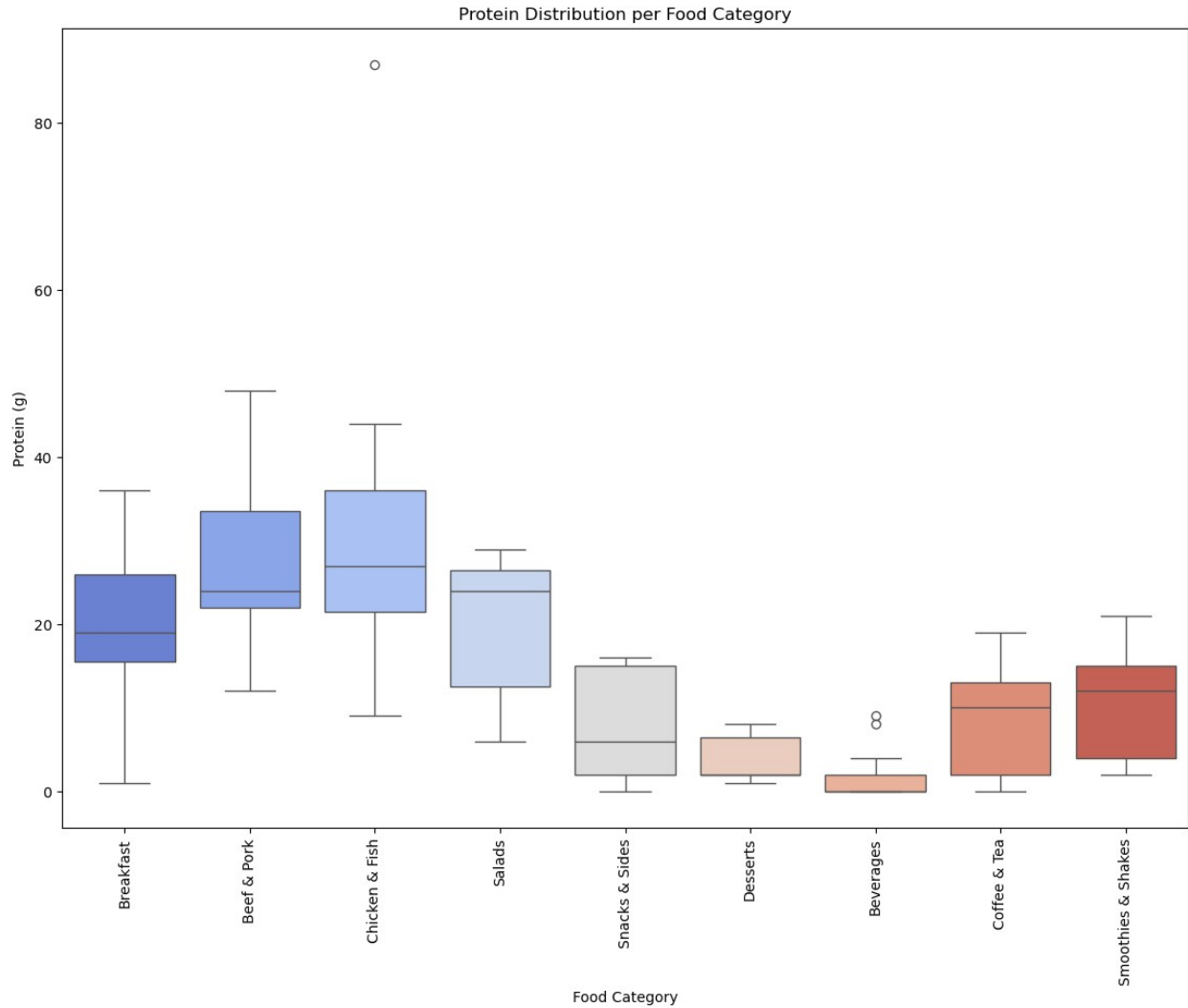
fig, ax = plt.subplots(figsize=(10, 4))
sns.barplot(x='Total Fat', y='Item', data=max_fat, ax=ax, hue='Item')

ax.set_title('Top 10 Items with Highest Total Fat Content')
ax.set_xlabel('Total Fat (g)')
ax.set_ylabel('Item')

plt.show()
```



```
# Box plot for protein distribution per category
plt.figure(figsize=(14, 10))
sns.boxplot(x='Category', y='Protein', data=df, hue='Category',
palette='coolwarm', dodge=False, legend=False)
plt.xticks(rotation=90)
plt.title('Protein Distribution per Food Category')
plt.xlabel('Food Category')
plt.ylabel('Protein (g)')
plt.show()
```

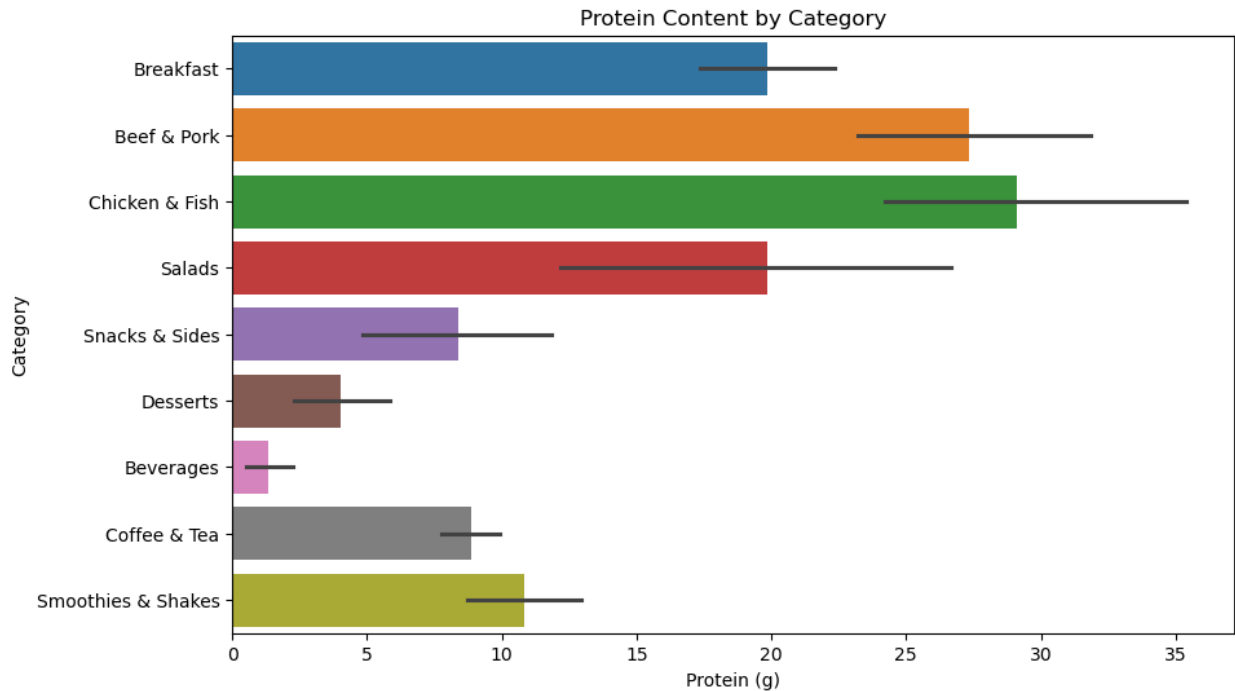


```
# Protein Content by Category
fig, ax = plt.subplots(figsize=(10, 6))

sns.barplot(x='Protein', y='Category', data=df, ax=ax, hue='Category')

ax.set_title('Protein Content by Category')
ax.set_xlabel('Protein (g)')
ax.set_ylabel('Category')

plt.show()
```

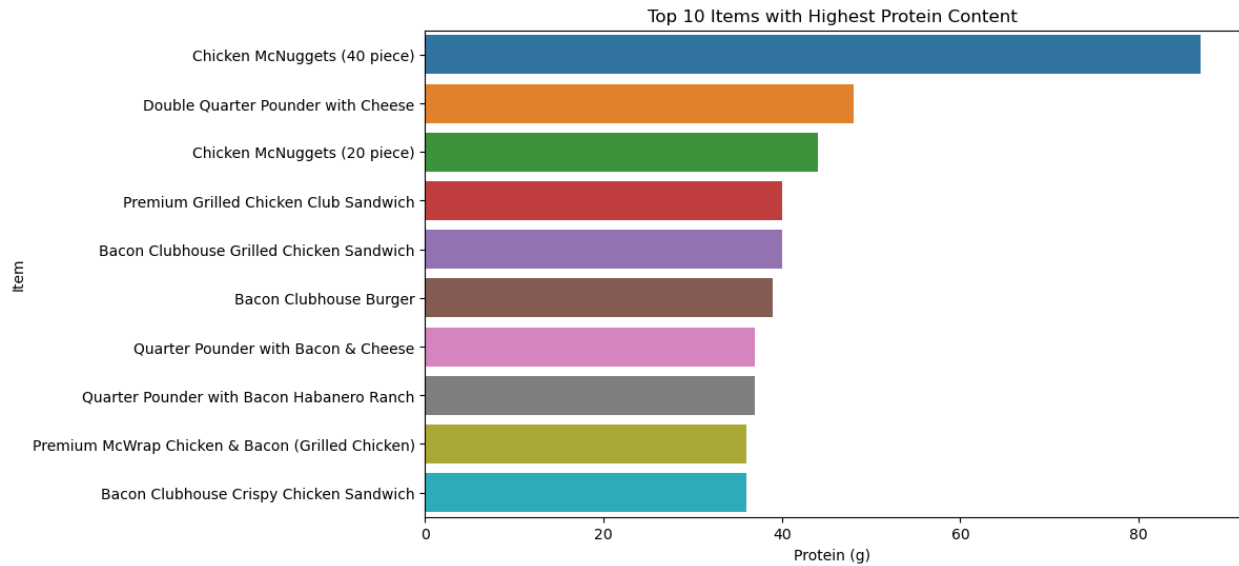


```
# Top 10 items with the highest Protein content
max_protein = df.sort_values('Protein', ascending=False).head(10)

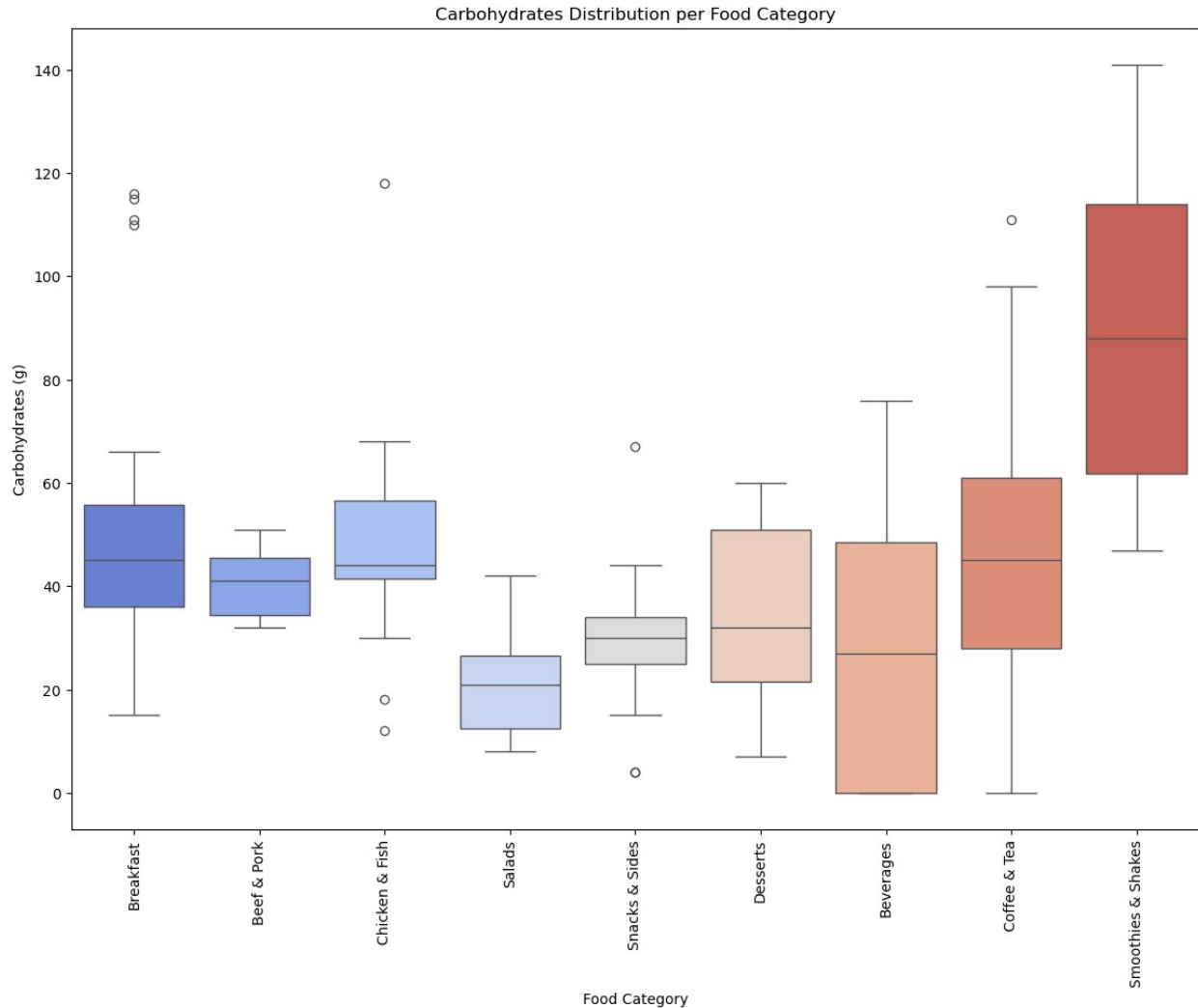
fig, ax = plt.subplots(figsize=(10, 6))

sns.barplot(x='Protein', y='Item', data=max_protein, ax=ax,
hue='Item')

ax.set_title('Top 10 Items with Highest Protein Content')
ax.set_xlabel('Protein (g)')
ax.set_ylabel('Item')
plt.show()
```



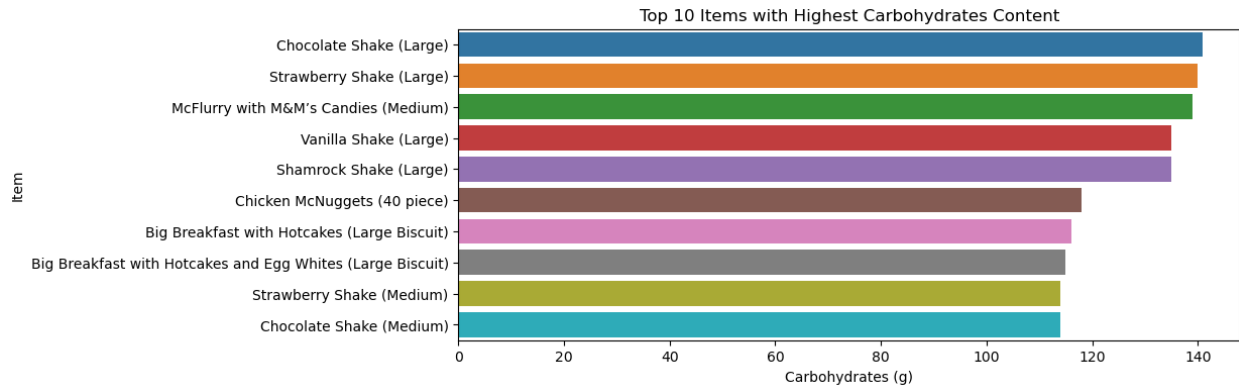
```
# Box plot for carbohydrate distribution per category
plt.figure(figsize=(14, 10))
sns.boxplot(x='Category', y='Carbohydrates', data=df, hue='Category',
palette='coolwarm', dodge=False, legend=False)
plt.xticks(rotation=90)
plt.title('Carbohydrates Distribution per Food Category')
plt.xlabel('Food Category')
plt.ylabel('Carbohydrates (g)')
plt.show()
```



```
# The top 10 items with the highest Carbohydrates content
max_carb = df.sort_values('Carbohydrates', ascending=False).head(10)

fig, ax = plt.subplots(figsize=(10, 4))
sns.barplot(x='Carbohydrates', y='Item', data=max_carb, ax=ax,
hue='Item')

ax.set_title('Top 10 Items with Highest Carbohydrates Content')
ax.set_xlabel('Carbohydrates (g)')
ax.set_ylabel('Item')
plt.show()
```

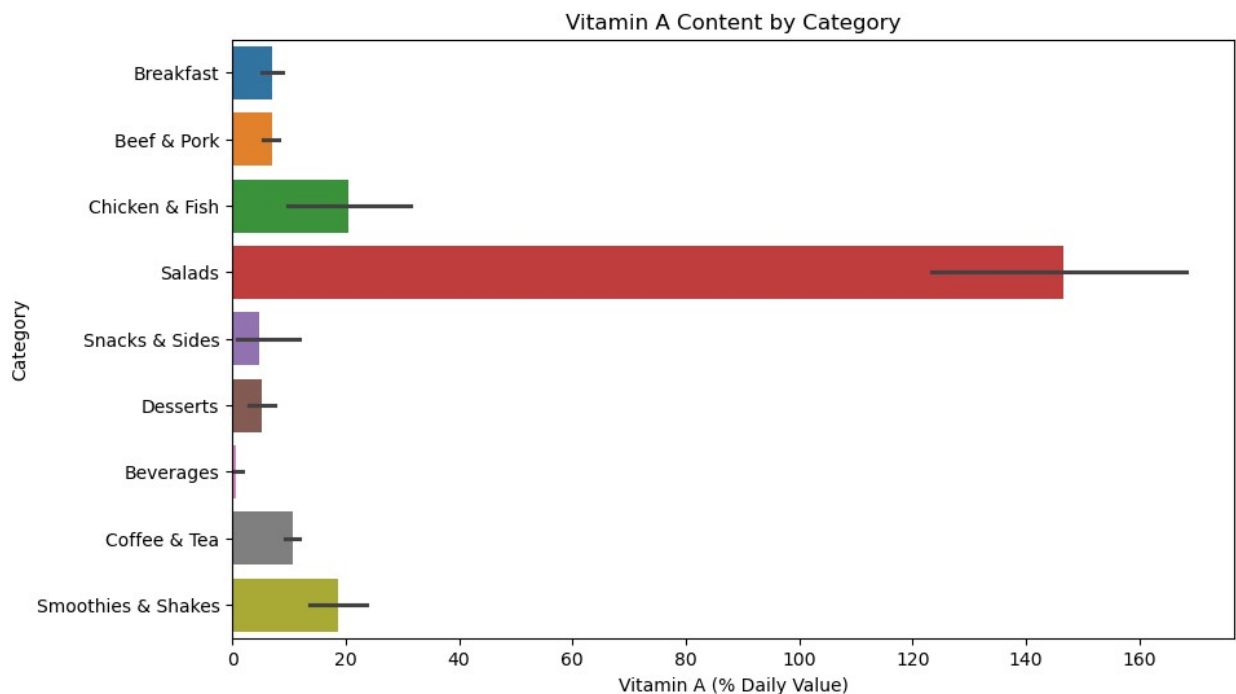



```
# Vitamin A Content by Category
fig, ax = plt.subplots(figsize=(10, 6))

sns.barplot(x='Vitamin A (% Daily Value)', y='Category', data=df,
ax=ax, hue='Category')

ax.set_title('Vitamin A Content by Category')
ax.set_xlabel('Vitamin A (% Daily Value)')
ax.set_ylabel('Category')

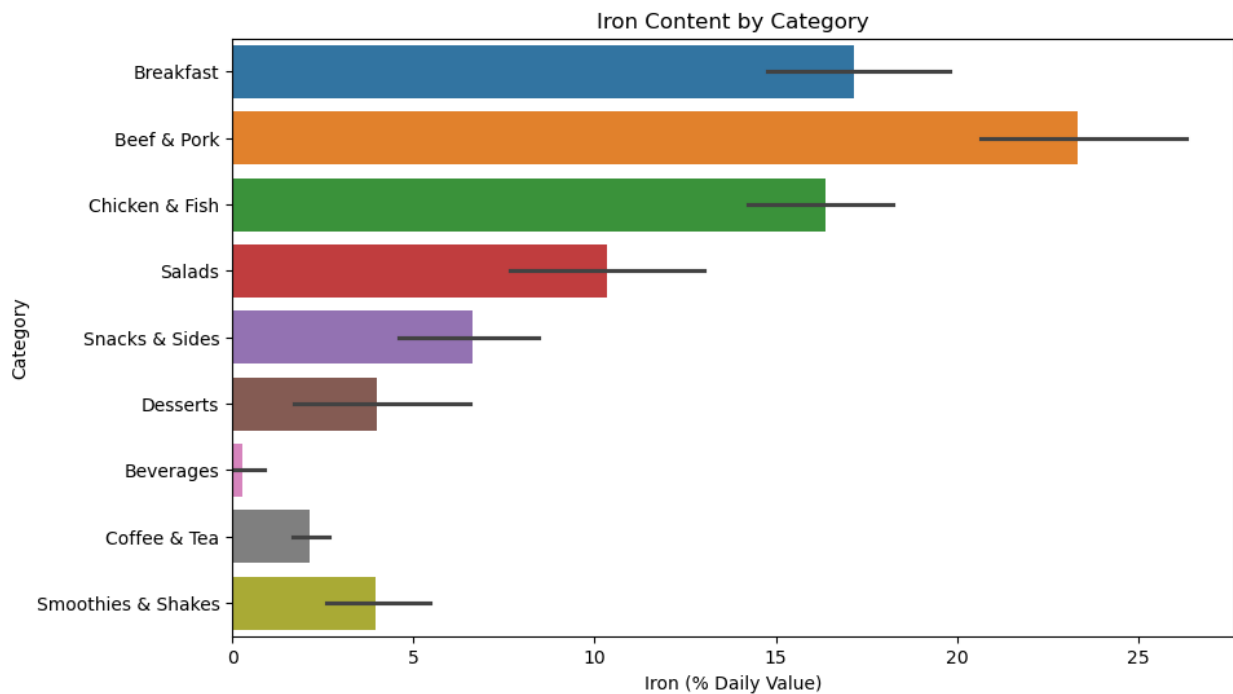
plt.show()
```



```
# Iron Content by Category
fig, ax = plt.subplots(figsize=(10, 6))
```

```
sns.barplot(x='Iron (% Daily Value)', y='Category', data=df, ax=ax,
hue='Category')
```

```
ax.set_title('Iron Content by Category')
ax.set_xlabel('Iron (% Daily Value)')
ax.set_ylabel('Category')
plt.show()
```



```
# Cholesterol Content by Category
```

```
fig, ax = plt.subplots(figsize=(15, 6))
```

```
sns.boxplot(x='Category', y='Cholesterol', data=df, ax=ax,
hue='Category')
```

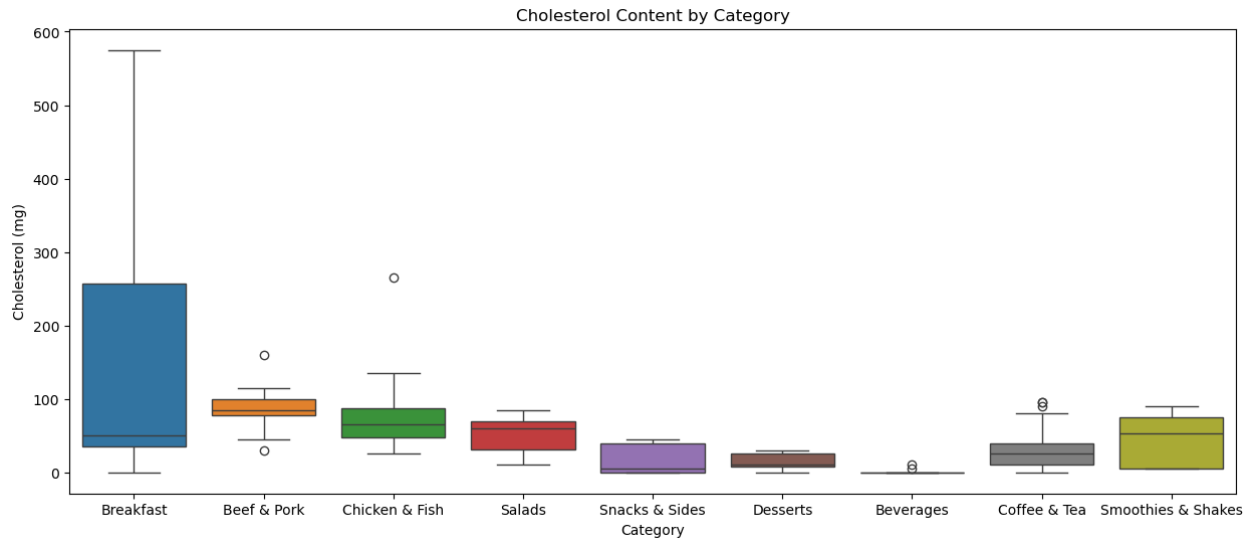
```
# Add a title and labels
```

```
ax.set_title('Cholesterol Content by Category')
```

```
ax.set_xlabel('Category')
```

```
ax.set_ylabel('Cholesterol (mg)')
```

```
plt.show()
```



Step 4: Nutrition-Based Insights

4.1 Identify menu items with the highest and lowest calorie counts:

```
# Highest calorie items
```

```
df.nlargest(10, 'Calories')[['Item', 'Category', 'Calories']]
```

	Item	
Category \		
82	Chicken McNuggets (40 piece)	Chicken & Fish
32	Big Breakfast with Hotcakes (Large Biscuit)	Breakfast
31	Big Breakfast with Hotcakes (Regular Biscuit)	Breakfast
34	Big Breakfast with Hotcakes and Egg Whites (La...	Breakfast
33	Big Breakfast with Hotcakes and Egg Whites (Re...	Breakfast
81	Chicken McNuggets (20 piece)	Chicken & Fish
253	McFlurry with M&M's Candies (Medium)	Smoothies & Shakes
246	Strawberry Shake (Large)	Smoothies & Shakes
249	Chocolate Shake (Large)	Smoothies & Shakes
243	Vanilla Shake (Large)	Smoothies & Shakes
	Calories	
82	1880	
32	1150	

31	1090
34	1050
33	990
81	940
253	930
246	850
249	850
243	820

Lowest calorie items

```
df.nsmallest(10, 'Calories')[['Item', 'Category', 'Calories']]
```

	Item	Category	Calories
114	Diet Coke (Small)	Beverages	0
115	Diet Coke (Medium)	Beverages	0
116	Diet Coke (Large)	Beverages	0
117	Diet Coke (Child)	Beverages	0
122	Diet Dr Pepper (Small)	Beverages	0
123	Diet Dr Pepper (Medium)	Beverages	0
124	Diet Dr Pepper (Large)	Beverages	0
125	Diet Dr Pepper (Child)	Beverages	0
136	Dasani Water Bottle	Beverages	0
137	Iced Tea (Small)	Coffee & Tea	0

4.2 Determine the average nutritional content of popular menu categories:

Average nutritional content per category

```
numeric_columns = df.select_dtypes(include='number').columns
```

```
numeric_df = df[numeric_columns]
```

```
df.groupby('Category')[numeric_columns].mean().round(2)
```

	Calories	Calories from Fat	Total Fat \
Category			
Beef & Pork	494.00	224.67	24.87
Beverages	113.70	0.74	0.09
Breakfast	526.67	248.93	27.69
Chicken & Fish	552.96	242.22	26.96
Coffee & Tea	283.89	71.11	8.02
Desserts	222.14	64.29	7.36
Salads	270.00	108.33	11.75
Smoothies & Shakes	531.43	127.68	14.12
Snacks & Sides	245.77	94.62	10.54

	Total Fat (% Daily Value)	Saturated Fat \
Category		
Beef & Pork	38.60	10.47
Beverages	0.15	0.06
Breakfast	42.67	10.65
Chicken & Fish	41.33	6.17

Coffee & Tea	12.36	4.92
Desserts	11.14	4.29
Salads	18.33	3.75
Smoothies & Shakes	21.71	8.38
Snacks & Sides	16.23	2.69

	Saturated Fat (% Daily Value)	Trans Fat
Cholesterol \ Category		

Beef & Pork	52.00	1.10
87.33		
Beverages	0.30	0.00
0.56		
Breakfast	53.43	0.11
152.86		
Chicken & Fish	31.11	0.13
75.37		
Coffee & Tea	24.37	0.14
27.26		
Desserts	21.29	0.00
15.00		
Salads	18.50	0.00
51.67		
Smoothies & Shakes	41.79	0.54
45.00		
Snacks & Sides	13.38	0.00
18.46		

	Cholesterol (% Daily Value)	Sodium	...
Carbohydrates \ Category			...

Beef & Pork	28.93	1020.67	...
40.13			
Beverages	0.19	41.48	...
28.81			
Breakfast	50.95	1211.07	...
49.76			
Chicken & Fish	25.22	1257.78	...
49.07			
Coffee & Tea	9.38	136.89	...
44.53			
Desserts	4.86	117.14	...
34.86			
Salads	17.33	588.33	...
21.67			
Smoothies & Shakes	14.71	183.57	...
90.43			
Snacks & Sides	6.23	395.77	...

29.15

	Carbohydrates (% Daily Value)	Dietary Fiber \
Category		
Beef & Pork	13.47	2.53
Beverages	9.59	0.04
Breakfast	16.57	3.26
Chicken & Fish	16.33	2.93
Coffee & Tea	14.86	0.78
Desserts	11.57	1.00
Salads	7.17	4.50
Smoothies & Shakes	30.14	1.46
Snacks & Sides	9.62	1.54

	Dietary Fiber (% Daily Value)	Sugars	Protein \
Category			
Beef & Pork	9.87	8.80	27.33
Beverages	0.07	27.85	1.33
Breakfast	12.83	8.26	19.86
Chicken & Fish	11.81	7.33	29.11
Coffee & Tea	3.18	39.61	8.86
Desserts	3.43	26.14	4.00
Salads	18.50	6.83	19.83
Smoothies & Shakes	5.75	77.89	10.86
Snacks & Sides	7.08	4.08	8.38

	Vitamin A (% Daily Value)	Vitamin C (% Daily
Value) \		
Category		
Beef & Pork	6.93	
7.33		
Beverages	0.74	
23.48		
Breakfast	6.93	
8.90		
Chicken & Fish	20.44	
12.63		
Coffee & Tea	10.74	
0.00		
Desserts	5.14	
4.14		
Salads	146.67	
28.33		
Smoothies & Shakes	18.75	
6.96		
Snacks & Sides	4.85	
28.15		

Calcium (% Daily Value) Iron (% Daily Value)

Category		
Beef & Pork	23.00	23.33
Beverages	2.96	0.30
Breakfast	16.17	17.14
Chicken & Fish	15.44	16.37
Coffee & Tea	28.29	2.15
Desserts	11.43	4.00
Salads	15.00	10.33
Smoothies & Shakes	35.57	3.96
Snacks & Sides	6.00	6.62

[9 rows x 21 columns]

Analysis and Reporting

Summary of Findings and Insights from the Analysis

1. Calorie Distribution:

- The majority of McDonald's menu items have calorie counts concentrated within a specific range, with several items being significant outliers.
- Desserts and burgers tend to have higher calorie counts compared to beverages and salads.

2. Nutritional Content Analysis:

- Total Fat: Burgers and breakfast items tend to have higher total fat content. Salads and beverages generally have lower fat content.
- Protein: Protein content is highest in burgers and chicken items, which makes them good sources of protein.
- Carbohydrates: Desserts and beverages have the highest carbohydrate content, primarily due to their high sugar content.
- Sugars: Desserts and beverages contain the most sugars, which aligns with their high carbohydrate content.
- Sodium: High sodium content is prevalent in burgers, chicken items, and breakfast options.

3. Correlation Analysis:

- There is a strong positive correlation between calories and total fat, indicating that higher-calorie items tend to have higher fat content.
- Calories also show a positive correlation with protein and sodium, suggesting that high-calorie items are typically richer in these nutrients as well.
- Carbohydrates and sugars are strongly correlated, which is expected as sugars contribute significantly to the carbohydrate content.

4. Category-wise Trends:

- Burgers: High in calories, total fat, protein, and sodium. They are energy-dense but also rich in nutrients that need moderation, such as fat and sodium.
- Salads: Generally lower in calories and fat but can vary significantly based on dressings and add-ons.
- Beverages: Wide range of calories, with sugary beverages contributing high sugar and carbohydrate content.
- Desserts: High in sugar and carbohydrates, moderate in calories, and low in protein and fat.
- Breakfast Items: High in calories, total fat, and sodium, with moderate protein content.

Benefits of Nutritional Analysis for McDonald's Customers and the Organization

Benefits for Customers

1. **Informed Choices:**
 - Nutritional transparency allows customers to make informed decisions about their food choices.
 - By understanding the nutritional content of each menu item, customers can choose meals that align with their dietary goals and health needs.
1. **Healthier Alternatives:**
 - With clear information, customers can identify healthier options.
 - For instance, they might opt for salads or grilled chicken items over higher-calorie burgers and fried foods.
1. **Dietary Management:**
 - Customers with specific dietary requirements, such as low sodium or low sugar diets, can use this information to select appropriate menu items, aiding in better health management.
1. **Portion Control:**
 - Knowing the calorie content of menu items can help customers practice portion control, thereby avoiding excessive calorie intake.

Benefits for the Organization

1. **Enhanced Customer Trust:**
 - Providing detailed nutritional information enhances transparency and builds trust with customers, showing McDonald's commitment to their health and well-being.
1. **Market Differentiation:**
 - In an increasingly health-conscious market, offering detailed nutritional information can differentiate McDonald's from competitors and attract health-conscious consumers.
1. **Menu Optimization:**

- Nutritional analysis can help McDonald's identify areas for menu improvement. For example, reducing sodium content in high-sodium items or offering lower-calorie versions of popular menu items can cater to health-conscious consumers.
1. **Targeted Marketing:**
 - Understanding the nutritional profiles of menu items allows McDonald's to better target their marketing efforts. For instance, they can promote items that align with popular dietary trends, such as high-protein or low-carb diets.
 1. **Regulatory Compliance:**
 - Detailed nutritional information ensures compliance with regulatory requirements regarding food labeling and nutritional transparency, avoiding potential legal issues and enhancing corporate responsibility.
 1. **Customer Feedback and Innovation:**

Analyzing customer preferences based on nutritional data can provide insights into consumer trends, enabling McDonald's to innovate and introduce new items that meet the evolving demands of their customer base.

Recommendations for McDonald's

1. **Introduce Healthier Options:**
 - Expand the menu to include more items that are low in calories, fat, and sodium. For example, offer more salads, grilled options, and fruit-based desserts.
1. **Nutritional Labeling:**
 - Ensure that all menu items have clear and accessible nutritional labeling, both in-store and online, to help customers make informed choices.
1. **Reformulate Recipes:**
 - Consider reformulating high-sodium, high-fat, and high-sugar items to reduce these components without compromising on taste.
1. **Portion Sizes:**
 - Offer smaller portion sizes for high-calorie items to provide customers with more choices and control over their calorie intake.
1. **Customer Education:**
 - Launch campaigns to educate customers about the importance of balanced nutrition and how to make healthier choices from the McDonald's menu.
1. **Sustainability and Sourcing:**
 - Emphasize sustainability and healthier sourcing practices, such as using organic ingredients, reducing additives, and ensuring high-quality nutritional standards.

By implementing these recommendations, McDonald's can improve the nutritional profile of its menu, cater to the growing demand for healthier food options, and enhance its brand image as a responsible and customer-focused organization.